Focus

I sistemi robotici ad autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?

Daniele Amoroso e Guglielmo Tamburrini*

INCREASINGLY AUTONOMOUS ROBOTIC SYSTEMS BETWEEN ETHICS AND LAW: WHAT ROLE FOR HUMAN CONTROL?

ABSTRACT: To be counted as operationally autonomous relative to the execution of some given task, a robotic system must be capable of performing that task without any human intervention after its activation. Recent progress in the fields of robotics and AI has paved the way to robots autonomously performing tasks that may significantly affect individual and collective interests, which are deemed as worthy of protection from both ethical and legal perspectives. The present contribution provides an overview of ensuing normative problems and identifies some ethically and legally grounded solutions to them. To this end, three case studies will be more closely scrutinized, i.e. increasingly autonomous weapons systems, vehicles, and surgical robots. These case studies are used to illustrate, respectively, the preliminary problem of whether we want to grant certain forms of autonomy to robotic systems, the problem of selecting appropriate ethical policies to control the behavior of autonomous robotic systems, and the problem of how to retain responsibility for misdoings of autonomous robotic systems. The analysis of these case studies brings out the key role played by human control in ethical and legal problem-solving strategies concerning the operational autonomy of robotic and AI systems.

KEYWORDS: autonomous weapons systems; self-driving cars; surgical robots; deontological ethics and consequentialism; international law

* *Daniele Amoroso: Professore associato di Diritto Internazionale, Dipartimento di Giurisprudenza, Università degli Studi di Cagliari. Mail: daniele.amoroso@unica.it. Guglielmo Tamburrini: Professore ordinario di Logica e Filosofia della scienza, Dipartimento di Ingegneria elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II. Mail: guglielmo.tamburrini@unina.it.* *Benché il presente lavoro sia il frutto di una riflessione congiunta dei due Autori, i paragrafi 3 e 4 sono da attribuire a Daniele Amoroso ed i restanti a Guglielmo Tamburrini. Il lavoro di ricerca sui temi trattati in questo articolo è stato parzialmente sostenuto dal progetto PRIN 2015_TM24JS. Contributo sottoposto al referaggio del Comitato Scientifico.*

Focus

1. Introduzione

L'espressione "sistema robotico autonomo" si riferisce genericamente a quei sistemi robotici che, una volta attivati, sono in grado di svolgere determinati compiti senza alcun ulteriore intervento da parte dell'operatore/utente umano. La rapida progressione che negli ultimi anni ha caratterizzato la ricerca in ambito robotico[1] ha reso tecnicamente possibile il compimento, da parte di tali sistemi, di attività suscettibili di incidere in modo significativo su posizioni individuali o collettive giuridicamente protette, o comunque considerate meritevoli di tutela sulla base di un giudizio etico. I casi più noti sono rappresentati dall'uso della forza letale da parte dei cd. sistemi d'arma autonomi e dalla circolazione dei veicoli autonomi, ma gli esempi potrebbero essere numerosi: si pensi ai robot chirurgici e a vari sistemi robotici per l'assistenza.[2] Questo stato di cose ha creato le condizioni per un forte rilancio della riflessione sulle implicazioni etico-giuridiche della robotica e dell'intelligenza artificiale (IA), che si è spinta ben oltre i circoli accademici e specialistici, entrando nel dibattito politico[3] e ricevendo una notevole copertura mediatica.[4]

Senza disconoscere la specificità dei problemi posti da ciascuna tecnologia, è possibile individuare alcune questioni che, seppure con diversa intensità, interessano trasversalmente i sistemi robotici autonomi. Anzitutto, vi è il discorso relativo all'opportunità – se non addirittura all'obbligo – di sostituire l'operatore/utente umano con macchine autonome, quando le prestazioni di queste ultime garantiscano una migliore protezione degli interessi in gioco (ad es. riducendo gli incidenti in strada, in sala operatoria o – *mutatis mutandis* – sul campo di battaglia). In una prospettiva opposta, va poi richiamata la discussione sul (presunto) diritto fondamentale a che decisioni suscettibili di incidere su posizioni individuali di particolare pregnanza sul piano etico-giuridico (*in primis*, il diritto alla vita e all'integrità fisica) siano prese da un essere umano e non da un agente robotico. Infine, nei casi in cui la sostituzione uomo-macchina sia ritenuta accettabile, occorre stabilire come regolamentare il comportamento dei sistemi robotici di fronte a problemi eticamente e giuridicamente complessi e come incardinare un rapporto di responsabilità in caso di eventi dannosi causati dalla macchina.

---

[1] B. SICILIANO e O. KHATIB (a cura di), *Handbook of Robotics*, Berlino-Heidelberg, II ed., 2016.

[2] Su quest'ultima, che non sarà oggetto di analisi, v. M. DECKER, *Caregiving robots and ethical reflection: the perspective of interdisciplinary technology assessment*, in *AI & Society*, 22(3), 2008, 315-330.

[3] Si vedano, solo per fare alcuni esempi, i lavori della Commissione sull'Intelligenza artificiale della *House of Lords* britannica, consultabili su: www.parliament.uk/ai-committee (ultima consultazione 7/02/2019); la Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)); il Parere "Sviluppi della Robotica e della Roboetica" del 17 luglio 2017, elaborato da un gruppo di lavoro misto costituito dal Comitato Nazionale per la Bioetica e dal Comitato Nazionale per la Biosicurezza, le Biotecnologie e la Scienza della Vita, entrambi istituiti presso la Presidenza del Consiglio dei Ministri italiana (reperibile al seguente indirizzo: http://presidenza.governo.it/biotecnologie/documenti/Robotica-misto%20CNB-CNBBSV-17lug17-IT.pdf (ultima consultazione 7/02/2019)).

[4] Per farsene un'idea è sufficiente consultare i siti web dei quotidiani italiani più popolari. V., ad esempio, *La guida autonoma alla conquista della fiducia dei pedoni*, in www.repubblica.it, 26 gennaio 2019 (ultima consultazione 7/02/2019); G. CORBELLINI, Se il robot è meglio del medico, in www.ilsole24ore.com, 9 luglio 2018 (ultima consultazione 7/02/2019); M. GAGGI, *La guerra dei robot: armi senza soldati e killer automatici*, in www.corriere.it, 15 aprile 2018 (ultima consultazione 7/02/2019).

Per orientarsi in questa discussione, appare utile ricordare sinteticamente le due principali famiglie di teorie dell'etica normativa, quella consequenzialista e quella deontologica, anche in ragione dell'indubbia influenza da esse esercitata nell'analisi delle questioni rilevanti per la roboetica (e quindi per un diritto della robotica eticamente orientato). Il consequenzialismo valuta la moralità di una determinata scelta esclusivamente alla luce della bontà delle conseguenze che ne derivano. Si inseriscono in questo solco, ad esempio, gli argomenti che promuovono l'introduzione di forme avanzate di autonomia in ragione delle migliori *prestazioni* raggiunte dai sistemi robotici a salvaguardia del benessere degli utenti. L'etica deontologica si focalizza, invece, sui doveri morali come guida per l'azione e come parametro di giudizio del valore morale delle scelte individuali e collettive. Si pensi, a questo proposito, all'assunto secondo cui certe attività "sensibili" sotto il profilo etico-giuridico debbano necessariamente essere condotte da esseri umani.

È ben noto che i principi dell'etica deontologica e dell'etica consequenzialista possono occasionalmente entrare in conflitto tra loro. Come vedremo, il perseguimento del benessere collettivo potrebbe motivare, da una prospettiva consequenzialista, la concessione di una maggiore autonomia al sistema robotico, secondo modalità che confliggono, almeno *prima facie*, con l'imperativo deontologico di riservare agli esseri umani lo svolgimento di alcune attività eticamente sensibili. Sarebbe tuttavia un errore applicare meccanicamente tale distinzione al fine di aggiudicare in una direzione o nell'altra i problemi morali posti dai sistemi robotici autonomi. Per affrontare efficacemente questi problemi, infatti, è necessaria un'indagine approfondita sulle norme morali rilevanti, che ne offra un'interpretazione contestuale, immaginandone l'applicazione in scenari concreti. L'esito di questa indagine può ben consistere nel raggiungimento di soluzioni di compromesso, pervenendo per questa via a proposte articolate di risoluzione dei conflitti morali da incorporare in strumenti (o interpretazioni) giuridicamente vincolanti.

Su queste premesse, il presente scritto intende offrire una visione d'insieme (ancorché necessariamente parziale) delle problematiche sollevate dai sistemi robotici autonomi e suggerire alcune possibili soluzioni. A questo scopo, dopo un breve chiarimento sulla nozione di autonomia rilevante ai fini della nostra indagine (par. 2), ci si soffermerà su tre casi di studio: i sistemi d'arma autonomi (par. 3), i veicoli autonomi (par. 4) e i robot chirurgici (par. 5). La discussione di questi casi, infatti, ci consentirà di illustrare – rispettivamente – il tema dell'accettabilità etica e giuridica dell'autonomia dei sistemi robotici, il problema della definizione delle regole di comportamento del sistema autonomo e la questione della responsabilità per eventi dannosi causati da agenti artificiali. Essa farà emergere, inoltre, la centralità del controllo umano nelle strategie di soluzione dei problemi etico-giuridici nel campo della robotica e dell'IA (par. 6).

2. L'autonomia dei sistemi robotici come "autonomia operativa"

È bene sgombrare subito il campo da una possibile fonte di confusione derivante dall'uso prevalente che si fa della parola "autonomia" nel linguaggio giuridico e nella filosofia morale. Le definizioni di ciò che in quegli ambiti si dice "autonomia personale" non sono utili per delimitare i sistemi robotici autonomi, poiché si attribuisce tale autonomia solo a entità che sono consapevoli di sé e del mondo circostante, e delle quali si presuppone la capacità di essere liberi e di agire in base a intenzioni proprie.

Nessun dispositivo che si possa realisticamente prefigurare in base alle conoscenze scientifiche e ingegneristiche correnti soddisfa tutte le condizioni per essere considerato un'entità genuinamente autonoma in questa accezione. Se ciò possa accadere in un orizzonte futuro indeterminato, e cioè in un tempo futuro non prevedibile sulla base degli attuali sviluppi scientifici e tecnologici, è questione concettualmente rilevante, ma di nessun interesse ai fini del problema morale e giuridico che si pone *oggi*, e con ben diversa urgenza, in relazione ai sistemi autonomi già esistenti o tecnologicamente imminenti che considereremo nei prossimi paragrafi.

Come accennato nel paragrafo introduttivo, la nozione di autonomia che impiegheremo in questo scritto è quella di "autonomia operativa", vale a dire relativa allo svolgimento di un determinato compito. In questa prospettiva, pertanto, un sistema robotico può essere definito "autonomo" in base alla sua capacità di eseguire un dato compito senza richiedere alcun intervento da parte degli esseri umani. I dispositivi tecnologici che godono in questo senso di autonomia operativa formano una classe ampia ed eterogenea, poiché il repertorio dei compiti comprende tanto attività complesse, quali la navigazione stradale dei veicoli autonomi o l'esecuzione di un'operazione chirurgica, quanto compiti che consideriamo molto più semplici dalla prospettiva della loro eseguibilità da parte di agenti artificiali, come ad esempio il controllo dell'accensione e dello spegnimento di una caldaia da parte di un termostato.

Nella classe dei sistemi che godono di autonomia operativa, dunque, confluiscono sia dispositivi nati dall'integrazione di componenti funzionali dotati di capacità percettive, cognitive e di coordinamento sensorimotorio, che si collocano alla confluenza della ricerca avanzata nei settori dell'IA e della robotica, sia sistemi semplici basati su tecnologie meno complesse e già sviluppate da tempo. Se è vero che l'indicazione di esempi paradigmatici dell'una e dell'altra categoria è piuttosto agevole, molto più difficile è individuare criteri generali e più precisi atti a distinguere, in modo univoco, la "parte alta" di questa classe da quella "bassa"[5] – una difficoltà, questa, che è figlia in ultima analisi della relativa vaghezza che caratterizza le descrizioni stesse dei domini d'indagine e degli scopi della robotica e dell'IA. Poco interessa in questa sede, tuttavia, il problema di demarcare con precisione alcune sottoclassi dell'ampia ed eterogenea classe dei sistemi robotici che godono di autonomia operativa. A rilevare, piuttosto, è la novità dei problemi etico-giuridici sollevati da alcuni di questi sistemi, la quale è legata allo svolgimento di compiti che richiedono capacità percettive, cognitive e di giudizio complesse, che fino a poco tempo fa erano di esclusiva pertinenza degli esseri umani. È chiaro, dunque, che stiamo guardando alla "parte alta" dell'insieme dei sistemi dotati di autonomia operativa, la cui realizzazione è resa oggi possibile *dalla confluenza delle tecnologie più avanzate dell'IA e della robotica*.

[5] Per una discussione delle varie impostazioni al problema dell'autonomia, v. il saggio di G. SARTOR e A. OMICINI, *The autonomy of technological systems and responsibilities for their use*, in N. BHUTA *et al.*, *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge, 2016, 39-74.

ISSN 2284-4503

Focus

3. L'(in)accettabilità etica e giuridica dell'autonomia nei sistemi robotici: il dibattito sulle armi autonome

Secondo una definizione elaborata dal Dipartimento della difesa statunitense, e che pare destinata a consolidarsi nella prassi internazionale, le armi autonome sono quelle armi che, una volta attivate, sono in grado di identificare, selezionare ed ingaggiare un obbiettivo senza ulteriore intervento umano.[6] Negli ultimi anni, la comunità internazionale – su impulso della società civile[7] – ha avviato un dibattito sulla loro legalità ed accettabilità etica, che ha visto coinvolti Stati, ONG ed organizzazioni internazionali, prevalentemente nella cornice istituzionale della cd. Convenzione sulle armi convenzionali[8], prima nell'ambito di incontri informali (2014-2016) e poi, a partire dal 2017, in seno ad un Gruppo di esperti governativi.[9]

Il motivo per cui questo sviluppo tecnologico ha creato tanto interesse (a differenza, ad esempio, di un drone che controlli autonomamente alcuni aspetti della propria navigazione aerea) sta nel fatto che, in questo caso, l'autonomia riguarda le funzioni *critiche* della selezione e dell'ingaggio degli obbiettivi nelle operazioni belliche. Tali funzioni sono considerate "critiche" perché la loro esecuzione *i)* è oggetto di approfondita regolamentazione nel diritto internazionale ed in particolare di quella branca del diritto internazionale dei conflitti armati (o diritto internazionale umanitario) che va sotto il nome di "*Law of targeting*"[10]; *ii)* è un fattore chiave ai fini della responsabilità individuale e statale; *iii)* implica scelte morali suscettibili di incidere, anche in misura assai profonda, su posizioni individuali eticamente rilevanti e giuridicamente tutelate (diritto alla vita ed all'integrità fisica, diritto ad una casa, e così via).

La discussione sull'accettabilità etica e giuridica di questa tecnologia militare riflette in larga misura quanto si è detto nel paragrafo introduttivo – in termini più generali – in riferimento ai sistemi robo-

[6] Dipartimento della Difesa statunitense, *Directive 3000.09: Autonomy in Weapon Systems*, 21 novembre 2012, 13-14. Benché la questione sia ancora controversa, è significativo che intorno alla definizione proposta dagli Stati Uniti (vale a dire: la potenza militare tecnologicamente più avanzata nel campo dell'intelligenza artificiale) vi sia una chiara convergenza di vedute da parte di altri importanti *stakeholders*, quali il Comitato internazionale per la croce rossa (*Autonomous weapon systems: Implications of increasing autonomy in the critical functions of weapons*, Ginevra, 2016, pp. 11-12) e le ONG che sostengono la campagna per la messa al bando delle armi autonome (v. Campaign to Stop Killer Robots, *Urgent Action Needed to Ban Fully Autonomous Weapons*, 23 aprile 2013, reperibile al seguente indirizzo: http://stopkillerrobots.org/wp-content/uploads/2013/04/KRC_LaunchStatement_23Apr2013.pdf (ultima consultazione 7/02/2019)).

[7] Il dibattito pubblico sulle implicazioni etico-giuridiche di questa tecnologia militare è stato stimolato dalla campagna "Stop Killer Robots", lanciata nell'aprile del 2013 da una coalizione internazionale di ONG con l'obiettivo dichiarato di promuovere l'adozione di un trattato che vieti lo sviluppo, la produzione, il possesso e l'uso dei sistemi d'arma autonomi. Per una cronologia dettagliata delle attività della Campagna, v. https://www.stopkillerrobots.org/action-and-achievements/ (ultima consultazione 7/02/2019).

[8] Convenzione sul divieto o la limitazione dell'impiego di talune armi classiche che possono essere ritenute capaci di causare effetti traumatici eccessivi o di colpire in modo indiscriminato, Ginevra, 10 ottobre 1980.

[9] La documentazione relativa agli incontri informali ed ai lavori del Gruppo di esperti governativi è consultabile su: https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument (ultima consultazione 7/02/2019).

[10] Sul quale v. M.N. SCHMITT e E. WIDMAR, *The Law of Targeting*, in P.A.L. DUCHEINE, M.N. SCHMITT e F.P.B. OSINGA (a cura di), *Targeting: The Challenges of Modern Warfare*, L'Aja, 2016, 121-145.

ISSN 2284-4503

tici autonomi. Gli argomenti contrari all'autonomia nei sistemi d'arma sono stati avanzati tanto in una prospettiva deontologica quanto in quella consequenzialista. Quelli favorevoli, invece, sono inquadrabili prevalentemente in termini consequenzialisti.

Gli argomenti deontologici impiegati per promuovere la messa al bando delle armi autonome sono tre. *Il primo argomento* sostiene che le armi autonome non sarebbero in grado di assicurare il rispetto dei principi cardine del diritto internazionale umanitario, vale a dire il principio di distinzione, il principio di proporzionalità ed il principio di precauzione. Il *principio di distinzione* impone alle parti belligeranti di dirigere i propri attacchi esclusivamente contro combattenti nemici ed obbiettivi militari (risparmiando, dunque, la popolazione e gli oggetti civili).[11] Una norma di contenuto analogo, ancorché non riconducibile formalmente nell'alveo del principio di distinzione, è quella che fa divieto di rivolgere gli attacchi contro nemici che abbiano manifestato l'intenzione di arrendersi o che siano stati messi fuori combattimento (*hors de combat*).[12] Il *principio di proporzionalità*, invece, proibisce di lanciare un attacco "da cui ci si può attendere che provochi incidentalmente morti e feriti fra la popolazione civile, danni ai beni civili, o una combinazione di perdite umane e danni [cd. danni collaterali], che risulterebbero eccessivi rispetto al vantaggio militare concreto e diretto previsto".[13] Il *principio di precauzione*, infine, è strumentale alla realizzazione dei primi due principi, in quanto richiede ai belligeranti di adottare tutte le misure praticabili al fine di evitare che un attacco sia diretto contro la popolazione e gli oggetti civili o che comunque provochi un danno collaterale sproporzionato.[14]

La possibilità che un'arma autonoma riesca a rispettare i principi di distinzione e proporzionalità in modo comparabile ad un combattente umano competente e coscienzioso presuppone la soluzione di numerosi e profondi problemi che attengono alla ricerca nei campi dell'IA e della robotica avanzata.[15] La qualificazione di un combattente nemico come *hors de combat*, ad esempio, implica la capacità di riconoscere comportamenti che veicolano messaggi di resa, intenzioni ingannevoli o la sopravvenuta inabilità a combattere.[16] Ancor più complesso, poi, è l'accertamento relativo alla "partecipazione diretta alle ostilità", in virtù della quale un civile può divenire un obbiettivo legittimo, e ciò anche a causa dell'incertezza e caoticità che caratterizzano gli scenari di guerriglia urbana. Il principio di proporzionalità, d'altro canto, pone la diversa ed ulteriore difficoltà – il cui superamento nel breve e medio periodo appare poco realistico[17] – di tradurre in un algoritmo la delicata attività di bilanciamento tra vantaggio militare anticipato e danno collaterale atteso. Infine, è legittimo dubitare che l'eliminazione della supervisione umana sulle attività di *targeting* svolte dalle armi autonome sia

[11] Primo Protocollo Addizionale alle Convenzioni di Ginevra del 12 agosto 1949 relativo alla protezione delle vittime dei conflitti armati internazionali, Ginevra, 8 giugno 1977 (Primo Protocollo Addizionale alle Convenzioni di Ginevra), Artt. 48, 51 co. 2 e 52 co. 2.

[12] Primo Protocollo Addizionale alle Convenzioni di Ginevra, Art. 41 co. 2.

[13] Primo Protocollo Addizionale alle Convenzioni di Ginevra, Art. 51 co. 5 lett. b).

[14] Primo Protocollo Addizionale alle Convenzioni di Ginevra, Art. 57.

[15] Questo aspetto è riconosciuto anche dai roboticisti in principio favorevoli allo sviluppo delle armi autonome, v. R. ARKIN, *Lethal Autonomous Systems and the Plight of the Non-combatant*, in AISB Quarterly, 137, 2013, 1-9, 4.

[16] R. SPARROW, *Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender*, in *International Law Studies*, 91, 2015, 699-728.

[17] N.E. SHARKEY, *The evitability of autonomous robot warfare*, in *International Review of the Red Cross*, 94, 2012, 787-799, 789-790.

compatibile con l'obbligo di adottare tutte le precauzioni possibili per evitare danni (sproporzionati) alla popolazione civile, tenuto conto del fatto che le tecnologie di IA – anche quelle più avanzate – sono soggette a errori imprevedibili, che un operatore umano sarebbe tuttavia in grado di evitare agevolmente.[18]

Il secondo argomento deontologico pone l'attenzione sul rischio che l'autonomia dei sistemi d'arma dia luogo a "vuoti di responsabilità" in caso di scelte di *targeting* materialmente contrarie al diritto internazionale. Anche i sostenitori più convinti dell'accettabilità etica e giuridica delle armi autonome sono infatti costretti ad ammettere che un'arma autonoma, per quanto sofisticata, possa erroneamente dirigere un attacco contro la popolazione civile o provocare danni collaterali sproporzionati, vale a dire atti che, se compiuti da un essere umano, sarebbero qualificabili come crimini di guerra. Chi ne risponderà? La lista dei potenziali responsabili include: il comandante militare che ha deciso l'operazione, l'operatore che ha attivato l'arma e tutti coloro che sono stati coinvolti nella realizzazione di quest'ultima (produttori, ingegneri robotici, programmatori) e nel relativo processo di *legal review*. Sennonché, la complessità di queste tecnologie e l'imprevedibilità del loro comportamento[19] sono suscettibili di escludere l'imputabilità di tutti i soggetti inclusi nella lista per difetto dell'elemento soggettivo (*mens rea*) dell'illecito, così come disciplinato dal diritto internazionale (penale). Questo risultato si pone in aperto contrasto con il dovere morale dei combattenti di rendere conto delle proprie azioni ed omissioni e del corrispondente principio di responsabilità penale individuale nel diritto internazionale.[20]

Il terzo ed ultimo argomento deontologico afferma l'incompatibilità con il principio della dignità umana dell'attribuzione ad un agente artificiale del compito di operare scelte morali riguardanti la vita, l'integrità fisica ed i beni delle persone coinvolte in un conflitto armato (o in altro contesto in cui l'arma può essere impiegata). Queste ultime, infatti, si troverebbero in una condizione di soggezione nei confronti di un decisore robotico che non ne condivide la condizione umana, e dunque private della possibilità di fare appello all'umanità di *qualcuno* che si trovi dall'altra parte. Il loro valore intrinseco in quanto esseri umani verrebbe così sistematicamente negato e violato.[21]

La portata degli argomenti sopra esposti contro l'autonomia dei sistemi d'arma è limitata agli usi letali e a quelli che possono avere un impatto negativo sulla popolazione e sugli oggetti civili. Sono quindi concepibili ipotesi di impiego delle armi autonome che si collocano in uno spazio non regolato

[18] Alcune ricerche sulle reti neurali profonde hanno dimostrato che è possibile indurre un programma per il riconoscimento automatico delle immagini a "vedere" uno struzzo nella foto di uno scuolabus, apportando a quest'ultima delle modifiche impercettibili per l'occhio umano. Per questo ed altri esempi, v. C. SZEGEDY *et al.*, *Intriguing properties of neural networks*, in *arxiv.org*, 19 febbraio 2014 (ultima consultazione 7/02/2019).

[19] Imprevedibilità che dipende non solo dalla complessità dell'IA installata nel sistema d'arma, ma anche dalle caratteristiche del contesto in cui essa è chiamata ad operare. V. G. TAMBURRINI, *On Banning Autonomous Weapon Systems: From Deontological to Wide Consequentialist Reasons*, in N. BHUTA *et al*., *op. cit.*, 122-141, 127-128.

[20] Su questo punto v., anche per ulteriori riferimenti, D. AMOROSO e B. GIORDANO, *Who Is to Blame for Autonomous Weapons Systems' Misdoings?*, in N. LAZZERINI e E. CARPANELLI (a cura di), *Use and Misuse of New Technologies. Contemporary Challenges in International and European Law*, L'Aja, 2019 (in corso di pubblicazione).

[21] P. ASARO, *On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making*, in *International Review of the Red Cross*, 94, 2012, 687-709, 689; C. HEYNS, *Autonomous weapons systems: living a dignified life and dying a dignified death*, in N. BHUTA *et al*., *op. cit.*, 3-19; A. SHARKEY, *Autonomous weapons systems, killer robots and human dignity*, in *Ethics and Information Technology*, 2018.

da principi deontologici. Si pensi, ad esempio, ad un attacco sferrato contro una struttura militare disabitata e sufficientemente separata da oggetti civili, o ad un duello aereo tra droni autonomi, senza esseri umani nel loro raggio di azione. In questi scenari, a ben vedere, non si pongono problemi di distinzione e proporzionalità, non c'è rischio che l'arma autonoma "commetta" un crimine di guerra, né che il principio della dignità umana venga violato.

Tale limite non trova applicazione quando si passano a considerare gli argomenti consequenzialisti, che includono sia argomenti favorevoli che contrari all'autonomia nei sistemi d'arma. Questo significa che un attacco contro una struttura militare disabitata, che può risultare non problematico in una prospettiva deontologica, può essere riprovevole da un punto di vista consequenzialista. Viceversa, determinati usi letali delle armi autonome, pur essendo inammissibili in un'ottica deontologica, potrebbero essere considerati eticamente desiderabili in ragione delle conseguenze che ne discendono.

È utile distinguere, a questo proposito, tra consequenzialismo ristretto e consequenzialismo ampio. *Il primo approccio* si focalizza sulle *prestazioni* delle armi autonome sul campo di battaglia e sulle conseguenze che ci si aspetta ne derivino nell'immediato. Questo approccio è alla base del principale argomento etico a favore delle armi autonome, secondo cui il loro sviluppo dovrebbe essere consentito in quanto esse porterebbero ad una riduzione nelle perdite umane in battaglia, sia tra i combattenti che tra i civili. Tale aspettativa si fonda sulla duplice convinzione *i)* che le armi autonome eseguiranno scelte di *targeting* più accurate di quelle umane e potranno essere programmate in modo da prendere decisioni meno aggressive, non essendo condizionate dall'istinto di autoconservazione[22] e *ii)* che l'impatto dello sviluppo delle armi autonome sarà limitato allo specifico contesto operativo in cui esse sono impiegate (assunto *ceteris paribus*). Quest'ultimo assunto è contestato dall'*approccio consequenzialista ampio*, il quale considera, oltre alle implicazioni sul campo di battaglia, anche gli effetti destabilizzanti connessi al dispiegamento delle armi autonome, che includono la neutralizzazione dei disincentivi a cominciare una guerra di aggressione – visto il coinvolgimento di un numero sempre minore di soldati –, la non prevedibilità delle interazioni con altre armi autonome, la loro vulnerabilità ad attacchi cibernetici e il conseguente rischio di guerre "accidentali", l'accelerazione dei conflitti a velocità che si pongono al di là delle capacità reattive e cognitive degli esseri umani, il rischio che tale tecnologia cada nelle mani di regimi oppressivi e gruppi terroristici, la corsa globale agli armamenti e le ricadute sulle politiche di deterrenza.[23]

Queste le linee principali del discorso sull'accettabilità etico-giuridica delle armi autonome, che forniscono altresì la cornice normativa di riferimento della discussione in corso a Ginevra nel quadro della Convenzione sulle armi convenzionali. A questo proposito va rilevato come, col passare degli anni, il dibattito in seno al Gruppo di esperti governativi si stia progressivamente focalizzando sul c.d. ele-

[22] R. ARKIN, *Governing Lethal Behavior in Autonomous Robots*, Boca Raton/Londra/New York, 2009, 29-36. Questo argomento è stato fatto proprio dalla delegazione statunitense nell'ambito della discussione in seno al Gruppo di esperti governativi. V. in particolare il *working paper* fatto circolare il 28 agosto 2018 (UN Doc. CCW/GGE.1/2018/WP.4).

[23] G. TAMBURRINI, *op. cit.*, 137-141; J. ALTMANN e F. SAUER, *Autonomous Weapon Systems and Strategic Stability*, in *Survival*, 59(5), 2017, 117-142. Analoghe preoccupazioni sono state manifestate nella Lettera aperta sottoscritta il 21 agosto 2017 da più di cento imprenditori attivi nel campo della robotica e dell'intelligenza artificiale (IA) (il testo della lettera è reperibile al seguente indirizzo: https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nations-ban-lethal-autonomous-weapons/).

mento umano, vale a dire sulla identificazione del tipo di interazione uomo-macchina preferibile sul piano etico-giuridico.[24] Il merito di aver chiaramente individuato nell'elemento umano la questione chiave nella riflessione sulle armi autonome va riconosciuto alla ONG britannica *Article 36* la quale, muovendo da posizioni sostanzialmente riconducibili al *secondo argomento deontologico* (i.e. quello fondato sulla responsabilità), ha posto l'accento sulla necessità di assicurare che gli attacchi sferrati da *tutti* i sistemi d'arma (inclusi, dunque, quelli autonomi) siano soggetti ad un "controllo umano significativo", laddove l'uso dell'aggettivo "significativo" è inteso ad escludere forme di controllo meramente nominale.[25]

Ma in cosa dovrebbe consistere il controllo dell'operatore? L'approccio più convincente, sotto questo profilo, è quello secondo cui il problema non si presterebbe ad una soluzione unica, valida per tutti i sistemi d'arma e per tutti gli scenari.[26] A ben vedere, infatti, una risposta giuridica articolata consegue naturalmente alla "trasversalità" della nozione di autonomia, la quale è suscettibile di essere applicata a sistemi d'arma con funzioni e proprietà molto diverse tra loro.

La futura discussione a Ginevra sul tema dell'elemento umano dovrebbe pertanto mirare a facilitare l'applicazione dei principi etico-giuridici analizzati in questo paragrafo in situazioni concrete, attraverso la formulazione di un adeguato insieme di *regole-ponte* che colleghino i primi alle seconde. Segnatamente, andrebbero individuati i fattori suscettibili di rendere problematico, sotto il profilo giuridico, lo svolgimento delle funzioni critiche della selezione e dell'ingaggio degli obbiettivi in sostituzione totale o parziale del decisore umano. Tra i fattori che potrebbero venire in rilevo a questo proposito possiamo ricordare: gli obbiettivi operativi assegnati all'arma (offensivi/difensivi), la cornice geografica e temporale nella quale essa è chiamata a funzionare in autonomia, le caratteristiche dello scenario operativo (e.g. presenza/assenza di civili), la tipologia di obbiettivi da ingaggiare (persone, oggetti "abitati", oggetti "disabitati"). Per ciascuna combinazione dei fattori così individuati, occorrerà poi definire il livello di controllo umano normativamente richiesto. Si potrebbe pensare ai seguenti livelli di interazione tra operatore ed arma a controllo umano decrescente:

L0: La scelta dell'obbiettivo da colpire è integralmente effettuata dall'operatore umano
L1: La scelta dell'obbiettivo da colpire è effettuata dall'operatore umano sulla base di un ventaglio di opzioni suggerite dal sistema
L2: L'operatore umano approva o rifiuta le scelte del sistema in merito agli obbiettivi da colpire
L3: L'operatore umano supervisiona le scelte di effettuate dal sistema, conservando la possibilità di riprendere il controllo ed annullare l'attacco

[24] *Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, 23 ottobre 2018, par. 28 (UN Doc. CCW/GGE.1/2018/3).
[25] Article 36, *Killer Robots: UK Government Policy on Fully Autonomous Weapons*, 19 aprile 2013, 3-4, reperibile al seguente indirizzo: http://www.article36.org/weapons-review/killer-robots-uk-government-policy-on-fully-autonomous-weapons-2/ (ultima consultazione 7/02/2019).
[26] Si veda, in proposito, la posizione del governo statunitense, che preferisce parlare a questo riguardo di «*appropriate levels of human judgment*» (UN Doc. CCW/GGE.2/2018/WP.4, parr. 8-15).

L4: L'operatore umano si limita ad attivare il sistema, definendone la missione, ma senza avere la possibilità di intervenire in seguito.[27]

Questi livelli saranno quindi considerati di volta in volta idonei oppure non idonei a garantire che il controllo umano continui a esercitare la propria funzione di salvaguardia, volta ad impedire che il cattivo funzionamento dell'arma si traduca in un attacco diretto contro la popolazione civile o comunque in danni collaterali sproporzionati, preservando un nesso tra azione umana ed evento dannoso tale da incardinare un rapporto di responsabilità giuridica e morale. È chiaro, dunque, che la presenza di civili, l'esistenza di un notevole iato temporale tra l'attivazione dell'arma e l'effettiva attuazione della missione o la definizione di un ambito operativo particolarmente ampio sotto il profilo spaziale, sono tutti fattori che spingeranno verso l'applicazione di un livello di controllo alto (L0, L1 o – se le circostanze non consentono diversamente – L2). L3 potrà rappresentare, d'altra parte, il livello di *default* per gli usi che, in assenza di tali fattori, appaiono meno problematici. La completa rinuncia al controllo umano nell'esecuzione della missione (L4), infine, dovrà considerarsi in linea di principio incompatibile con il quadro etico-giuridico qui delineato, ad esclusione delle ipotesi – da ritenersi eccezionali – in cui la supervisione da parte di un operatore umano non solo sia impraticabile, ma sia altresì suscettibile di mettere in pericolo la vita e l'incolumità delle persone. Si pensi, ad esempio, ad un sistema anti-missile la cui funzione è quella di reagire ad attacchi di saturazione, rispetto ai quali è necessaria una risposta immediata, potenzialmente incompatibile coi tempi di reazione umani.

4. Come disciplinare l'autonomia operativa in situazioni eticamente e giuridicamente complesse: veicoli autonomi e collisioni inevitabili

Una tassonomia delle modalità di interazione uomo-macchina per certi versi analoga a quella proposta nel precedente paragrafo è stata adottata dalla *Society of Automotive Engineers International* (SAE International), che ha introdotto una gerarchia di livelli di autonomia crescente (L0-L5) per gli autoveicoli.[28] La gerarchia culmina con il livello L5, nel quale si situano i veicoli autonomi in grado di svolgere *tutti* i sottocompiti nei quali è scomponibile l'attività della guida, in *ogni* parte della rete stradale e in *qualsiasi* condizione di luce e di meteo. I livelli inferiori della gerarchia individuano un percorso di graduale avvicinamento all'autonomia completa, che passa attraverso un numero crescente di sottocompiti svolti dall'autoveicolo indipendentemente dagli esseri umani, forme di condivisione del controllo della navigazione che assegnano alla macchina privilegi via via più ampi e l'eliminazione progressiva di limiti operativi (relativi, ad es., alla velocità, alle porzioni di rete stradale, etc.).

È interessante notare come la navigazione autonoma di un veicolo ai livelli più alti della gerarchia sarà resa possibile (o comunque agevolata) da una rete di agenti informatici, ciberfisici o robotici ai quali è collegato e dai quali può dipendere per eseguire correttamente il compito assegnato: pro-

[27] N.E. SHARKEY, *Staying in the loop: human supervisory control of weapons*, in N. BHUTA *et al*., *op. cit.*, 23-38, 34-37.

[28] SAE International, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (J3016), giugno 2018, 19.

grammi informatici che gli forniscono la mappa dinamica di altri veicoli collocati al di fuori del raggio d'azione dei suoi sensori, altri veicoli autonomi con i quali si coordina, sistemi di localizzazione GPS e segnali ricevuti da sensori fissi installati sulla rete stradale.[29] Questo aspetto segna un'importante differenza, *di carattere tecnico*, rispetto all'autonomia dei sistemi d'arma. Mentre questi ultimi sono destinati ad operare in *scenari ostili*, popolati da attori che mirano ad ostacolarne il corretto funzionamento, i veicoli autonomi sono pensati per navigare in *ambienti cooperativi*, che favoriscono la prevedibilità e dunque la controllabilità dei loro comportamenti.

Questa circostanza – unitamente ad altre, più ovvie (in particolare: la prospettiva di un uso prevalentemente pacifico di questa tecnologia) – fa sì che il dibattito intorno alla desiderabilità etica dei veicoli autonomi non sia caratterizzato dalle opposizioni di principio che abbiamo riscontrato, invece, in riferimento all'autonomia dei sistemi d'arma. Al contrario, vi sono forti ragioni etiche che spingono per lo sviluppo di veicoli ad autonomia crescente. L'argomento principale a supporto di tale tecnologia va individuato nell'aspettativa che essa porterà con sé una notevole diminuzione degli incidenti stradali, il cui verificarsi è in larga misura determinato da fattori umani: tempi di reazione inadeguati, errori di percezione, inesperienza, violazione intenzionale del codice della strada, abuso di alcol e droghe. Questa aspettativa offre una motivazione etica in favore dei veicoli autonomi tanto in un'ottica consequenzialista, in vista della massimizzazione del benessere collettivo, quanto in una deontologica, in virtù del dovere – gravante su tutti coloro che sono coinvolti nella circolazione stradale – di rispettare la vita e l'integrità fisica altrui.

Considerazioni di questo tenore (oltre che, naturalmente, i giustificati interessi economici dell'industria automobilistica e di altre parti interessate) sono alla base della scelta del legislatore tedesco di modificare la Legge sulla circolazione stradale (*Straßenverkehrsgesetz* (StVG)) allo scopo di autorizzare – quantomeno *in linea di principio* – la circolazione di veicoli autonomi dotati di autonomia L3 e L4 secondo la tassonomia SAE International.[30] Che tale autorizzazione sia stata concessa solo in linea di principio discende dal fatto che la navigazione autonoma dei veicoli è sì consentita, ma col vincolo del rispetto della disciplina prescritta dalle pertinenti norme di diritto internazionale e dell'Unione Europea, le quali – come si vedrà subito – sono decisamente meno permissive.[31]

Un primo ostacolo alla liberalizzazione dei veicoli autonomi a livello internazionale risiede nel fatto che i principali trattati in materia di circolazione stradale – vale a dire, la Convenzione di Ginevra del 1949[32] e la Convenzione di Vienna del 1968[33] – sono stati adottati in un'epoca storica in cui le ricerche nei campi della robotica e dell'IA erano agli albori e la navigazione autonoma era ben lontana dall'essere concepita come un progetto realistico. Di conseguenza, entrambe le Convenzioni prescri-

[29] Per questo motivo, sarebbe più corretto attribuire la capacità di navigazione autonoma al sistema tecnologico distribuito che si estende oltre la carrozzeria del veicolo stesso e che concorre con i suoi vari componenti a gestirne la navigazione. Tale aspetto è evidenziato dalla SAE International, che per questo motivo preferisce non parlare di veicolo "autonomo" (SAE International, *op. cit.*, 28).

[30] StVG, Sezione 1a (che richiama espressamente la classificazione SAE International).

[31] StVG, Sezione 1a, co. 3.

[32] Convenzione sul traffico stradale, Ginevra, 19 settembre 1949.

[33] Convenzione sul traffico stradale, Vienna, 8 novembre 1968.

vono che alla guida di un veicolo vi sia un "conducente"[34] e che questi sia una "persona" [35], stabilendo inoltre con un certo grado di dettaglio le funzioni del controllo umano.[36]

Vero è che, almeno in riferimento alla Convenzione di Vienna, tale limite è stato recentemente rimosso con l'aggiunta del comma 5bis all'art. 8, a norma del quale i sistemi di navigazione autonomi sono ammessi, a condizione che essi possano essere "neutralizzati" o "scavalcati" (cd. privilegi di *override*) o disattivati in qualsiasi momento dall'utente.[37] Sennonché, l'impatto concreto di questa novità è praticamente nullo per i Paesi membri dell'Unione Europea, in quanto quest'ultima non solo non è parte della Convenzione di Vienna, ma si è vincolata al rispetto delle ben più rigide regole formatesi nel quadro dell'Accordo UNECE[38] del 1958 sull'adozione di regolamenti tecnici armonizzati delle Nazioni Unite per i veicoli a ruote,[39] i cui regolamenti sono stati incorporati nella Direttiva quadro 2007/46/CE.[40]

Ai fini della nostra discussione, a venire in rilievo è il Regolamento n. 79 ("Disposizioni uniformi relative all'omologazione dei veicoli per quanto riguarda lo sterzo"), il quale – pur essendo stato modificato negli ultimi anni per tenere conto dello sviluppo di sistemi di navigazione autonoma[41] – conserva nei confronti di questi ultimi una posizione di netta chiusura. Le ragioni di tale approccio sono esplicitate nel paragrafo introduttivo del Regolamento, che richiama le "preoccupazioni legate all'attribuzione del controllo del veicolo" e l'"assenza di protocolli internazionali di trasmissione dati riguardo al controllo dello sterzo dall'esterno del veicolo". Le uniche concessioni all'autonomia dei veicoli riguar-

[34] Convenzione di Ginevra, art. 8, co. 1; Convenzione di Vienna, art. 8 co. 1.

[35] Convenzione di Ginevra, art. 4; Convenzione di Vienna, art. 1 lett. v). Questa precisazione impedisce di interpretare l'espressione "conducente" in chiave evolutiva, sì da includervi anche il sistema di guida autonoma, secondo l'approccio seguito – ma con riferimento alla normativa interna in tema di omologazione dei veicoli – dalla *National Highway Traffic Safety Administration* statunitense (Lettera interpretativa del 4 febbraio 2016, indirizzata a Chris Urmson, Direttore del Progetto "Self-Driving Car" di Google).

[36] V., per la particolare efficacia della formulazione, l'art. 13 co. 1 della Convenzione di Vienna («Ogni conducente di veicolo deve, in ogni circostanza, restare padrone del proprio veicolo, in modo da potersi conformare alle esigenze della prudenza e da essere costantemente in grado di effettuare tutte le manovre che gli competono. Deve, regolando la velocità del proprio veicolo, tenere costantemente conto delle circostanze, in particolare della disposizione dei luoghi, dello stato della strada, dello stato del carico del proprio veicolo, delle condizioni atmosferiche e dell'intensità della circolazione, in modo da poter arrestare il proprio veicolo nei limiti del proprio campo di visibilità verso l'avanti, nonché dinanzi ad ogni ostacolo prevedibile. Deve rallentare e, se necessario, fermarsi tutte le volte che le circostanze lo esigano, in particolare quando la visibilità non è buona.»). Con riferimento alla Convenzione di Ginevra, v. artt. 8 co. 5, 10, 11 e 12 co. 4.

[37] Emendamento approvato dagli Stati parte il 26 marzo 2014 ed entrato in vigore il 23 marzo 2016. Per l'argomento secondo cui i veicoli autonomi sarebbero compatibili con la Convenzione di Ginevra anche in assenza di un emendamento in tal senso v. B.W. SMITH, *Automated Vehicles Are Probably Legal in the United States*, in *Texas A&M Law Review*, 1, 2014, 411-521, 424-457. La rilevanza pratica di questa conclusione risiede nel fatto che gli Stati Uniti hanno ratificato la Convenzione di Ginevra, ma non quella di Vienna.

[38] L'acronimo sta per *United Nations Economic Commission for Europe*.

[39] Accordo concernente l'adozione di regolamenti tecnici armonizzati delle Nazioni Unite per i veicoli a ruote, gli equipaggiamenti e i pezzi che possono essere installati o usati in veicoli a ruote, nonché le condizioni per il riconoscimento reciproco di omologazioni concesse sulla base di tali regolamenti delle Nazioni Unite, Ginevra, 20 marzo 1958.

[40] Direttiva 2007/46/CE del Parlamento europeo e del Consiglio, del 5 settembre 2007 che istituisce un quadro per l'omologazione dei veicoli a motore e dei loro rimorchi, nonché dei sistemi, componenti ed entità tecniche destinati a tali veicoli, Art. 34.

[41] V. *infra* le note 42 e 43 ed il testo che le accompagna.

ISSN 2284-4503

dano, in particolare, *i)* lo svolgimento di "una singola manovra laterale (ad esempio un cambio di corsia) previo apposito comando da parte del conducente" (nel gergo del Regolamento: "funzione sterzante a comando automatico di categoria C")[42] e *ii)* la "funzione sterzante di emergenza", vale a dire il "comando in grado di rilevare automaticamente il rischio di una collisione e di attivare in modo automatico lo sterzo del veicolo, per un periodo limitato, al fine di evitare la collisione (o di ridurne gli effetti)" con un altro veicolo o un ostacolo che si trovi o stia per trovarsi in traiettoria.[43]

Anche negli spazi piuttosto ristretti concessi dal Regolamento n. 79 – e segnatamente nella possibilità di omologare veicoli dotati di funzione sferzante di emergenza – vi sono i margini per una discussione delle questioni etiche poste dai veicoli autonomi, con specifico riferimento a quello che è in breve tempo diventato un "classico" del dibattito su questo tema, vale a dire il dilemma delle collisioni inevitabili.[44] Questo problema, che costituisce una rivisitazione in chiave contemporanea (e per certi versi più realistica) del noto "dilemma del carrello"[45], può essere illustrato con un esempio. Si consideri uno scenario in cui un veicolo dotato di funzione sterzante di emergenza debba evitare due ciclisti in tandem che, a causa delle cattive condizioni del manto stradale, si vengano a trovare improvvisamente nella sua traiettoria. Si consideri poi che l'unica manovra che impedirebbe la collisione, quasi certamente mortale per i ciclisti, sia suscettibile di causare – con lo stesso grado di certezza – la morte di un pedone che sta passeggiando a bordo strada. Questa situazione fa emergere una tensione etica tra approccio consequenzialista e deontologico. Il primo, infatti, considererà moralmente desiderabile effettuare la manovra, la quale consentirebbe di minimizzare il danno causando una sola vittima invece di due; il secondo, invece, giudicherà lo stesso corso d'azione contrario al principio della dignità umana in quanto teso a strumentalizzare la vita di un essere umano per la salvezza di altri.[46] Una variante di questo scenario prevede che il veicolo debba scegliere tra la morte del passeggero e quella di due o più pedoni. In questo caso, oltre alla tensione etica già descritta, sorgerebbe l'ulteriore problema – molto più prosaico – dell'appetibilità commerciale di un veicolo programmato per uccidere l'utente al fine di salvare altre vite umane.

[42] Regola 2.3.4.1.4., introdotta il 10 ottobre 2017. Secondo la classificazione SAE International, l'autonomia di un veicolo dotato di questa funzione andrà qualificata come di livello L1 (SAE International, *op. cit.*, 19).

[43] Regola 2.3.4.3., introdotta il 16 ottobre 2018. È interessante notare come la dotazione di questa funzione, pur ponendo importanti questioni etiche e giuridiche (sulle quali v. subito appresso), non è di per sé idonea a qualificare un veicolo come autonomo in base alla classificazione SAE International, la quale presuppone lo svolgimento prolungato (*sustained*) dei sottocompiti che compongono l'attività di guida, e non un intervento momentaneo (SAE International, *op. cit.*, 2). Tale circostanza è indicativa del fatto che non sempre vi è perfetta coincidenza tra quel che conta come *autonomo* a fini ingegneristici e quello che invece rileva in una prospettiva normativa.

[44] P. LIN, *Why Ethics Matters for Autonomous Cars*, in M. MAURER *et al.* (a cura di), *Autonomes Fahren*, Berlino, 2015, 69-85. V., anche per ulteriori riferimenti, P. SOMMAGGIO e E. MARCHIORI, *Break the chains: a new way to consider machine's moral problems*, in questa *Rivista*, 2018, 241-257.

[45] Sul quale v. D. EDMONDS, *Uccideresti l'uomo grasso? Il dilemma etico del male minore* (trad. it. di G. Guerriero), Milano, 2014.

[46] Per una potente affermazione di questo principio deontologico, v. la celebre sentenza del Tribunale Costituzionale federale tedesco che ha sancito l'incostituzionalità, per violazione del principio della dignità umana di cui all'art. 1 co. 1 del *Grundgesetz*, della previsione della legge tedesca sulla sicurezza aerea del 2005 che consentiva l'abbattimento, da parte delle forze armate, di un aereo civile (con passeggeri e personale di bordo) dirottato da parte dei terroristi al fine di colpire obiettivi civili o militari (BVerfG, *Urteil des Ersten Senats*, 15 febbraio 2006, 1 BvR 357/05).

Pur non affrontando espressamente il problema nella prospettiva delle collisioni inevitabili, il Regolamento n. 79 offre una disciplina che sembra rispondere ad un approccio deontologico rigorosamente incentrato sul rispetto delle norme sulla circolazione stradale. A norma del Regolamento, in particolare, le manovre effettuate da un veicolo con funzione sterzante di emergenza *i)* "non devono portare il veicolo a uscire dalla carreggiata";[47] *ii)* non devono comportare il superamento della segnaletica orizzontale[48] o comunque uno scartamento laterale del veicolo superiore a 0,75 metri;[49] *iii)* e soprattutto non devono "provocare una collisione del veicolo con altri utenti della strada".[50] È evidente come una disciplina di questo tenore (ed in particolare l'ultima delle regole richiamate ponga vincoli molti restrittivi per considerazioni di carattere consequenzialista sulla minimizzazione del danno.

Una limitata apertura in questo senso è invece rinvenibile nel Rapporto della Commissione Etica su "Guida automatica e connessa", nominata dal Ministero dei Trasporti e delle Infrastrutture Digitali tedesco, che presenta il pregio di affrontare il tema sulla base di un robusto impianto filosofico.[51] Il codice etico elaborato dalla Commissione prevede una linea guida specifica in materia di collisioni inevitabili che combina, invero in misura ineguale, preoccupazioni di carattere deontologico e consequenzialista.[52] Anzitutto, viene accordata priorità assoluta ai divieti – entrambi di marca prettamente deontologica – di operare discriminazioni in base alle caratteristiche personali (età, genere, condizione fisica e mentale) delle persone potenzialmente coinvolte nell'incidente e di fare calcoli compensativi sul numero delle vittime. Fatto salvo il rispetto di questi divieti, viene consentito – questa volta in un'ottica consequenzialista – di programmare i veicoli autonomi in modo da limitare i danni, purché il danno sia equamente ripartito tra tutti coloro che sono coinvolti nella collisione.[53]

Pur divergendo nei contenuti, la disciplina prevista nel Regolamento n. 79 e quella auspicata dalla Commissione etica tedesca sono entrambe espressione di un approccio di tipo paternalistico, volto ad imporre *dall'alto* le soluzioni ai dilemmi morali suscitati dal problema delle collisioni inevitabili, interferendo così con le preferenze etiche degli utenti. Queste ultime sono invece valorizzate dall'approccio libertario, in virtù del quale gli utenti possono decidere come la propria vettura dovrà comportarsi in caso di collisioni inevitabili. Seguendo tale approccio, alcuni autori hanno ipotizzato l'introduzione di una "manopola etica" attraverso la quale l'utente può stabilire *ex ante* se il veicolo autonomo dovrà *i)* sacrificare in ogni caso la vita del passeggero per salvare quella degli altri (modalità altruistica); *ii)* optare per la manovra che assicuri il minor numero di vittime, includendo eventualmente il passeggero (modalità imparziale); *iii)* accordare preferenza alla vita del passeggero (mo-

[47] Regola 5.1.6.2.3.
[48] Regola 5.1.6.2.3.1.
[49] Regola 5.1.6.2.3.2.
[50] Regola 5.1.6.2.4.
[51] Il Rapporto è stato tradotto in inglese ed è reperibile sul sito web del Ministero: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.html (ultima consultazione 7/02/2019).
[52] Si tratta della Regola 9 la quale nella traduzione inglese recita: «In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited. It is also prohibited to offset victims against one another. General programming to reduce the number of personal injuries may be justifiable. Those parties involved in the generation of mobility risks must not sacrifice non-involved parties».
[53] Per una discussione delle valutazioni etiche poste alla base di questa Regola, v. 17-19 del Rapporto.

dalità egoistica).[54] In questo modo, i veicoli si limiterebbero ad attuare in concreto le scelte etiche stabilite in termini generali dall'utente, agendo pertanto come "delegati morali" (*moral proxies*) di quest'ultimo.[55] Il comportamento del veicolo in situazioni eticamente complesse, dunque, rimarrebbe sottoposto al controllo umano, ancorché mediato dall'attività sensoriale e valutativa di una macchina. Sul piano giuridico, la riconducibilità del comportamento della macchina ad una scelta consapevole dell'utente consente di individuare in quest'ultimo il soggetto responsabile per i danni cagionati "intenzionalmente" (*rectius*: non determinati da un malfunzionamento) dalla vettura. Al tempo stesso, tuttavia, nella misura in cui i danni, inclusa la perdita di vite umane, siano stati provocati al fine di salvare la vita del passeggero (modalità egoistica) o altrui (modalità imparziale), l'utente potrà invocare a sua difesa – in presenza delle altre condizioni eventualmente previste dal diritto nazionale applicabile – la causa di giustificazione dello stato di necessità.[56]

5. Autonomia delle macchine e responsabilità (professionale) umana: il caso dei robot chirurgici

Il nesso tra controllo umano e responsabilità sta emergendo anche nel dibattito sui robot chirurgici. È indicativo, a riguardo, che nel fissare una gerarchia di sei livelli di autonomia per tali sistemi robotici, Yang et al. abbiano allo stesso tempo evidenziato – con una terminologia che ricorda da vicino la discussione sulle armi autonome – che i chirurghi continueranno a conservare a lungo una posizione di controllo *significativa*.[57] La gerarchia suggerita da questi autori colloca al livello più basso i sistemi che si limitano ad eseguire i comandi dell'operatore umano, costituendone una mera estensione robotica (L0); ai livelli successivi troviamo gli assistenti robotici che condizionano o correggono l'azione umana (L1), i sistemi robotici che eseguono compiti designati dall'operatore umano sotto la sua supervisione (L2) e i robot che generano strategie di esecuzione dei compiti sotto la supervisione umana (L3). A completare la classificazione vi sono i sistemi robotici, il cui sviluppo si situa in un futuro tecnologicamente distante, in grado di effettuare un'intera procedura con o senza supervisione umana (rispettivamente L4 e L5).

Nel campo della chirurgia robotica, i sistemi ad autonomia L0 sono impiegati come dispositivi *slave* per garantire elevata precisione nei micromovimenti ed attenuare eventuali tremori. Un tipico esempio di sistema tele-operato ad autonomia L0 è offerto dall'unità chirurgica laparoscopica "da Vinci", dove i chirurghi esercitano un controllo diretto sull'intera procedura, inclusa l'analisi dei dati, la

[54] G. CONTISSA, F. LAGIOIA e G. SARTOR, *The ethical knob: ethically customisable automated vehicles and the law*, in *Artificial Intelligence and Law*, 25, 2017, 365–378.

[55] J. MILLAR, *Technology as moral proxy: Autonomy and paternalism by design*, in *ETHICS 2014. Proceedings of the IEEE 2014 International Symposium on Ethics in Engineering, Science, and Technology*, 2014, 1-7.

[56] CONTISSA, LAGIOIA e SARTOR, *op. cit.*, 368-369. Nel senso che lo stato di necessità potrebbe essere efficacemente invocato in un numero molto limitato di circostanze, v. F. SANTONI DE SIO, *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*, in *Ethical Theory and Moral Practice*, 20, 2017, 411-429, la cui analisi tuttavia è prevalentemente incentrata sull'esperienza giuridica anglo-americana.

[57] G.-Z. YANG *et al.*, *Medical robotics – regulatory, ethical, and legal considerations for increasing levels of autonomy*, in *Science Robotics*, 2017, 2(4), 1-2, 2.

pianificazione pre- ed intra-operativa, le decisioni e l'effettiva esecuzione.[58] È chiaro che un sistema di questo tipo non pone particolari problemi dal punto di vista della significatività del controllo umano. Problemi più sottili emergono, invece, se passiamo a considerare i livelli di autonomia L1-L3.[59]
Diversi robot chirurgici impiegati in sala operatoria hanno già raggiunto autonomia L1. Un caso significativo, in questo senso, è costituito dai sistemi robotici che assistono i chirurghi nello spostamento del manipolatore lungo i percorsi di lavoro desiderati o impediscono ai manipolatori robotici di entrare in determinate aree di lavoro.[60] I sistemi robotici che identificano e applicano questi vincoli attivi (cd. *Virtual Fixtures*) non sono semplici dispositivi *slave*, poiché a volte correggono i movimenti decisi dal chirurgo. Per esercitare un controllo significativo a questo livello di autonomia, è necessario avere la possibilità di scavalcare le correzioni robotiche, mediante privilegi di controllo umano di secondo livello che consentano al chirurgo di prevalere sulle correzioni robotiche di primo livello.
Al livello di autonomia L2, gli umani selezionano un compito che i robot chirurgici devono eseguire. Il ruolo di supervisione del chirurgo consiste nel monitoraggio a mani libere e nella possibile "neutralizzazione" dell'esecuzione robotica. Il sistema robotico è dunque sotto il controllo *discreto* (piuttosto che *continuo*) del chirurgo. Un esempio di sistema dotato di autonomia L2 da tempo in uso nelle sale operatorie è ROBODOC, il quale esegue piani preoperatori relativi al taglio di ossa (*bone-milling*) sotto supervisione umana.[61] Un prototipo di ricerca più recente, pure classificabile come L2, è la piattaforma sperimentale Smart Tissue Autonomous Robot (STAR), che esegue la sutura intestinale (anastomosi) sul tessuto dei suini. Nei test sperimentali condotti su questo modello animale, il sistema STAR è risultato in grado di assicurare *prestazioni* migliori di quelle di chirurghi umani esperti.[62]
I sistemi chirurgici ROBODOC e STAR sono caratterizzati da diversi Livelli di Maturità Tecnologica (LMT). Il primo è utilizzato per procedure cliniche standard, mentre il secondo è ancora in fase di ricerca. Questa disparità dipende in modo cruciale dalla natura dei rispettivi ambienti operativi, e in particolare dalla loro prevedibilità. I siti chirurgici di ROBODOC sono strutture anatomiche rigide; il sistema STAR opera invece su tessuti morbidi deformabili. Gli ambienti strutturati in cui opera ROBODOC consentono di eseguire in sicurezza l'attività autonoma, grazie alla possibilità di effettuare misurazioni accurate e di prevedere gli eventi suscettibili di modificare la situazione. Di converso, i siti chirurgici morbidi e deformabili in cui verrà impiegato il sistema STAR sollevano sfide ben più ardue in relazione al rilevamento e tracciamento accurato, tanto degli strumenti chirurgici che delle parti anatomiche. Le differenze che intercorrono tra gli ambienti operativi di ROBODOC e STAR suggeriscono

[58] E. ACKERMAN, *New Da Vinci Xi Surgical Robot Is Optimized for Complex Procedures*, in *IEEE Spectrum*, 7 aprile 2014.
[59] F. FICUCIELLO, G. TAMBURRINI, A. AREZZO, L. VILLANI e B. SICILIANO, *Autonomy in surgical robots and its meaningful human control*, in *Paladyn Journal of Behavioural Robotics*, 10, 2019, 30-43.
[60] V. ad esempio ER2 (Eye Robot2), un robot per la microchirurgia oculare sviluppato dai ricercatori della Johns Hopkins University (A. ÜNERI *et al.*, *New steady-hand eye robot with micro-force sensing for vitreoretinal surgery*, in *2010 3rd IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2010, 814–819).
[61] N.A. NETRAVALI, M. BÖRNER e W.L. BARGAR, *The Use of ROBODOC in Total Hip and Knee Arthroplasty*, in L.E. RITACCO, F.E. MILANO e E. CHAO (a cura di), *Computer-Assisted Musculoskeletal Surgery Thinking and Executing in 3D*, Cham, 2016, 219-234.
[62] A. SHADEMAN *et al.*, *Supervised autonomous robotic soft tissue surgery*, in *Science Translational Medicine*, 8, 2016, 337–364.

di modulare opportunamente la vigilanza percettiva e cognitiva richiesta all'operatore affinché sia garantito un controllo umano significativo dei robot chirurgici che si collocano nell'ampia classe dei sistemi ad autonomia L2. Il campionamento percettivo episodico e la valutazione cognitiva dell'esecuzione dei compiti da parte del robot sono plausibilmente più impegnativi nel caso di sistemi come STAR, viste le possibili modifiche della situazione dovute al fisiologico flusso di sangue ed alla respirazione, e alla corrispondente necessità di valutare la risposta adattativa del robot. Pertanto, uno stesso tipo di controllo umano a carattere discontinuo non sarà necessariamente in grado di soddisfare il requisito della significatività in relazione a *tutti* i robot chirurgici della classe L2.

I robot chirurgici autonomi L3 generano strategie di esecuzione dei compiti sotto la supervisione umana, e fanno affidamento sull'operatore per scegliere tra le diverse strategie generate o per attuare una strategia selezionata autonomamente. In misura limitata, il sistema STAR raggiunge questo livello di autonomia per quanto riguarda la generazione di strategie di anastomosi[63], insieme ai sistemi che identificano dinamicamente le *virtual fixtures* e generano parametri o traiettorie di controllo ottimali.[64] Per essere significativo, il controllo umano sui robot L3 presuppone che i chirurghi decidano in modo competente se approvare o meno una delle strategie generate dal robot. Questa decisione implica che i chirurghi comprendano la logica delle soluzioni proposte, siano nella posizione di confrontare i rispettivi pro e contro e decidano in tempo utile quale strategia seguire. A seconda della complessità delle strategie proposte e dei siti chirurgici, il requisito del controllo umano significativo può porre sfide sempre più complesse concernenti l'interpretabilità da parte dell'operatore delle soluzioni generate dal robot e dunque la possibilità di prendere una decisione informata sulle stesse. Problemi simili possono emergere in relazione alle strategie che i robot chirurgici possono *imparare* a proporre sulla base di metodi di apprendimento automatico (cd. *machine learning*), in considerazione dei noti problemi di interpretabilità che possono influenzare questi sistemi, e in particolare quei sistemi che sono stati sviluppati in conformità con i metodi del cd. *deep learning*.[65] Oggi, l'apprendimento delle strategie chirurgiche è destinato ad essere basato su insiemi di dati formati da strategie generate dall'uomo. In un futuro più lontano, le questioni di interpretabilità che sorgono nel contesto del controllo umano sui robot chirurgici L3 potrebbero diventare sempre più acute se i dati impiegati per imparare a generare strategie di intervento riguardassero le strategie generate dai robot stessi ed i corrispondenti esiti clinici.

In sintesi, al fine di identificare correttamente il contenuto del requisito del controllo umano significativo, occorre anzitutto considerare le funzionalità che sono in grado di definire ed ordinare, in senso gerarchico, i livelli di autonomia dei robot chirurgici (il *cosa* dell'autonomia), gli ambienti corporei in cui questi sono chiamati ad operare (il *dove* dell'autonomia) e le capacità che il sistema impiega – ad es. l'apprendimento – per agire autonomamente (il *come* dell'autonomia).

[63] *Ibidem*.

[64] M. YIP e N. DAS, *Robot Autonomy for Surgery*, in R. PATEL (a cura di), *The Encyclopedia of Medical Robotics*, Singapore, 2018, 281-313.

[65] Su questo problema, e sui tentativi di venirne a capo, v. S. CHAKRABORTY *et al.*, *Interpretability of deep learning models: a survey of results*, relazione presentata nell'ambito dell'*IEEE Smart World Congress 2017*, 7-8 agosto 2017, il cui testo è reperibile al seguente indirizzo: http://orca.cf.ac.uk/101500/ (ultima consultazione 7/02/2019).

Da un punto di vista etico, l'applicazione di questo requisito ai robot chirurgici ad autonomia crescente è motivata dai principi bioetici di beneficenza e non-maleficenza, e dalle responsabilità deontologiche che ne derivano in capo ai chirurghi. È dunque necessaria un'analisi approfondita delle responsabilità *prospettiche* che verranno introdotte dal requisito del controllo umano significativo per delineare i programmi di formazione in chirurgia robotica. In particolare, il principio bioetico della non-maleficenza richiede una adeguata formazione volta a fornire gli strumenti concettuali che compensino i pregiudizi positivi nei confronti della macchina, che possono indurre erroneamente i chirurghi a fidarsi troppo del robot e troppo poco del proprio libero apprezzamento. Un'analisi approfondita dei compiti relativi al controllo umano significativo giocherà inoltre un ruolo altrettanto significativo nel valutare quali sono le (eventuali) responsabilità *retrospettive* del chirurgo, in caso di lesioni o morte del paziente operato mediante il supporto di robot chirurgici. Il corretto esercizio dei compiti discendenti dal requisito del controllo umano significativo potrà infatti costituire lo standard di diligenza alla luce del quale valutare eventuali profili di responsabilità (civile e penale). Con specifico riferimento all'ordinamento italiano, ad esempio, questo requisito potrebbe essere oggetto di "linee guida" o "buone pratiche clinico-assistenziali" il cui rispetto è suscettibile di escludere la punibilità del medico ai sensi dell'art. 590-sexies del codice penale[66] e può venire in rilievo come elemento a favore del medico nei giudizi di responsabilità civile.[67]

6. Conclusioni

L'analisi che precede ha evidenziato un elemento comune, che tiene insieme i tre casi di studio presi in considerazione e ne giustifica una trattazione congiunta, a dispetto delle palesi differenze che intercorrono tra l'uno e l'altro sistema robotico: la ricerca di una definizione delle forme e dei contenuti del controllo umano moralmente desiderabile (o, come si è detto, "significativo"), sulla cui base costruire una disciplina giuridica dei sistemi robotici eticamente orientata.

Portare a compimento tale ricerca costituisce un obbiettivo ambizioso, la cui realizzazione sarà possibile attraverso uno sforzo collettivo della comunità scientifica e degli altri attori interessati e che pertanto si pone ben al di là delle possibilità di questo breve scritto. In queste righe conclusive, quindi, ci limiteremo a fissare quelle che ci sembrano alcune coordinate di riferimento per le future ricerche su questo tema.

Per cominciare, è bene osservare come il problema del controllo umano significativo non si presti a soluzioni univoche. Ciò vale non solo – com'è ovvio che sia – nei rapporti tra le diverse classi di sistemi robotici (è intuitivo, ad esempio, che la questione si ponga in termini diversi per le armi autonome e per i robot chirurgici), *ma anche all'interno della medesima classe*. Il contenuto del requisito del controllo umano, infatti, va determinato alla luce dei compiti concretamente svolti dalla macchina, del contesto operativo e delle capacità impiegate a tal fine, vale a dire alla luce del *cosa*, del *dove* e

[66] Disposizione introdotta dalla Legge n. 24 dell'8 marzo 2017 (cd. legge "Gelli-Bianco"), il cui co. 2 recita: «Qualora l'evento si sia verificato a causa di imperizia, la punibilità è esclusa quando sono rispettate le raccomandazioni previste dalle linee guida come definite e pubblicate ai sensi di legge ovvero, in mancanza di queste, le buone pratiche clinico-assistenziali, sempre che le raccomandazioni previste dalle predette linee guida risultino adeguate alle specificità del caso concreto».

[67] Cassazione civile, sez. III, sentenza 9 maggio 2017, n. 11208, par. 1-a).

del *come* dell'autonomia di uno specifico sistema robotico. Questa distinzione è stata illustrata in riferimento ai robot chirurgici, ma può essere agevolmente applicata anche agli altri sistemi considerati. Per le armi autonome, ad esempio, il *cosa* riguarderà la funzione (difensiva/offensiva) e la tipologia di obbiettivi; il *dove* lo scenario operativo, avendo particolare riguardo alla presenza di civili; ed il *come*, le specifiche proprietà dell'arma (ad es., la capacità di sorvolare un'ampia porzione di territorio e di agire al suo interno per un periodo di tempo prolungato: cd. *loitering*).

Un altro aspetto sul quale appare opportuno richiamare l'attenzione riguarda le funzioni del controllo umano, che ne giustificano l'introduzione come requisito etico-giuridico, contribuendo altresì a definirne il contenuto. Il controllo umano opera in primo luogo come *meccanismo di salvaguardia* (*fail-safe mechanism*), volto ad impedire che il cattivo funzionamento della macchina provochi danni altrimenti evitabili. Si pensi a quanto detto sopra a proposito della compatibilità tra piena autonomia dei sistemi d'arma e principio di precauzione nel diritto internazionale umanitario; ma anche alla necessità di garantire all'operatore dei veicoli autonomi e dei robot chirurgici privilegi di *override* sulle scelte della macchina. In secondo luogo, il controllo umano costituisce un *catalizzatore di responsabilità*: in caso di eventi dannosi, infatti, esso consente di individuare il soggetto responsabile e le ragioni della riprovevolezza della sua condotta. L'esame dei doveri deontologici e delle connesse responsabilità professionali del chirurgo che operi servendosi di sistemi robotici appare particolarmente indicativo sotto questo profilo. In terzo luogo, il controllo umano assicura che sia un *agente morale*, e non una macchina, a prendere decisioni riguardanti la vita, l'integrità fisica e i beni delle persone coinvolte.[68] La caratterizzazione, secondo quello che abbiamo definito approccio libertario, dei veicoli autonomi come "delegati morali" o *moral proxies* dell'utente va precisamente in questa direzione.

Il controllo umano, tuttavia, non dev'essere considerato un feticcio, da difendere ad ogni costo. Motivazioni etiche, soprattutto di carattere consequenzialista, possono giustificare una forte limitazione del ruolo del decisore umano in ragione delle ricadute positive che l'autonomia di alcuni sistemi robotici può avere sul benessere collettivo e individuale. Si pensi, in proposito, alla riduzione degli incidenti stradali che ci si aspetta possa derivare dalla diffusione dei veicoli autonomi – un aspetto, questo, che non pare aver ricevuto sufficiente considerazione a livello internazionale, vista la rigidità che ha fino ad ora caratterizzato la normativa in materia. D'altra parte, nel formulare giudizi etici sui meriti dell'autonomia non ci si può limitare a considerarne le conseguenze immediate e dirette. La desiderabilità morale delle armi autonome, ad esempio, non può essere valutata solo sulla base degli eventuali benefici che potrebbero derivare ai combattenti ed alla popolazione civile *nello specifico scenario operativo in cui sono impiegate*, ma – come si è visto – anche alla luce degli effetti destabilizzanti che il loro sviluppo potrebbe determinare su larga scala.

[68] In senso analogo, v. F. SANTONI DE SIO e J. VAN DEN HOVEN, *Meaningful Human Control over Autonomous Systems: A Philosophical Account*, in F*rontiers in Robotic and AI*, 5-15, 2018, i quali pongono, tra le condizioni di significatività del controllo umano sui sistemi autonomi, la possibilità di istituire un nesso tra le attività di questi ultimi e (almeno) un essere umano (cd. *"tracing" condition*).